

Performance Comparison of Single Channel Speech Enhancement Techniques for Personal Communication

Milind U. Nemade¹, Prof. Satish K. Shah²

Associate Professor, Dept. of Electronics and Telecommunication, K. J. Somaiya Institute of Engineering and IT, Sion, Mumbai, India¹

Professor, Electrical Engineering Department, Faculty of Engineering and Technology, M. S. University of Baroda, Baroda, Gujrat, India²

ABSTRACT: Speech has been embedded into many applications like speech recognition, development of hearing aid, VoIP, mobile and other forms of personal communication. Speech enhancement techniques have been widely used for minimizing undesirable background noises. This paper deals with single channel speech enhancement techniques based on Spectral Subtraction (SS), Wavelet Transform (WT) and Adaptive Wiener Filtering (AWF). Here quantitative performance of these speech enhancement techniques is compared and the parameters used for comparison are Mean Square Error, Normalised Mean Square Error, Signal to Noise Ratio, Peak Signal to Noise Ratio and Average Absolute Distortion. The results obtained have proved the speech enhancing capability of the personal communication technique where noise and echo-interference can degrade the original speech signal. From the results we conclude that the performance of single channel speech enhancement based WT is better than AWF and SS techniques. MATLAB GUI developed for speech enhancement techniques which help to be able to visualize the results obtained throughout this paper.

Keywords: VoIP, Spectral Subtraction, Discrete Wavelet Transform (DWT), Adaptive Wiener Filter

I. INTRODUCTION

Speech is common mode of communication in the midst of human being and also the most efficient and reliable form of exchanging information among human. Speech has been embedded into many applications like speech recognition, development of hearing aid, VoIP, mobile and other forms of personal communication. In speech communication systems to get improvement in quality of speech at low bit rate, Linear Predictive Coding (LPC), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) are used [1]. In environment undesirable noise causes undesired effects in speech signal transmission and reception. Speech enhancement techniques have been widely used for minimizing these undesirable background noises. According to specific application, the requirement of speech enhancement technique varies to increase speech quality, intelligibility and performance of speech communication devices. VoIP (voice over internet protocol) played a vital role in communication system. Echo is the problem occurs in VoIP which reduces the quality of speech signal. It is difficult to remove echo completely but it can remove to tolerable range. If we try to remove it completely then it degrades the quality of speech signal on VoIP system [2]. Speech enhancer is required to improve the quality of degraded speech in VoIP system. Fig. 1 shows the required flow of process in VoIP system for echo cancellation and speech enhancement.

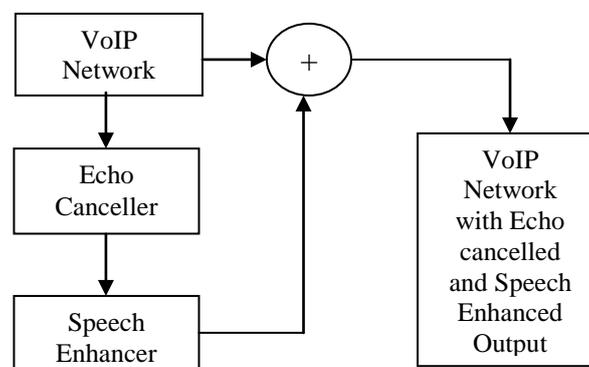


Fig.1 VoIP with Echo Canceller and Speech Enhancer

Speech enhancement techniques are broadly classified depending on the number of microphones used in system. Single channel speech enhancement system does not use reference noise but uses voice activity detector to characterized noise statistics during non voice region. Since last five decades, various approaches for noise reduction and speech enhancements have been investigated and developed. Single channel speech enhancement techniques mostly based on transform domain, adaptive filtering and model based techniques [3]. In this paper we focus on the quantitative performance comparison of transform domain and adaptive filtering single channel speech enhancement techniques for personal communication.

Recent advances in CPU and multi-core hardware has provided ample amount of computational power and thus, need for today is to design the complex but yet efficient and realistic approach for noise reduction to achieve speech enhancement. There are various types of advanced speech enhancement algorithms classified in main three categories, namely; filtering/estimation based noise reduction, beam forming and active noise cancellation (ANC) techniques [4]. In this paper, work has two-fold objectives. First is to implement Spectral Subtraction, Wavelet Transform and Adaptive Wiener filtering based single channel speech enhancement techniques using Graphical User Interface (GUI) in MATLAB for personal communication. Second, we attempted to fairly evaluate and compare the objective performance of these speech enhancement techniques using performance parameters like Signal to Noise Ratio, Peak Signal to Noise Ratio, Mean Square Error, Normalised Root Mean Square Error and Average Absolute Difference. Subjective performance is also compared along three dimensions SIG, BAK and OVRL.

The remaining part of the paper is organized as follows: In next section II, works related to the single channel speech enhancement techniques have been presented. Section III explains the objective evaluation of WT, SS and AWF speech enhancement techniques. Section IV explains subjective evaluation using three dimensions SIG, BAK and OVRL. Section V explains the MATLAB GUI implementation of above three speech enhancement techniques. Section VI specifies the results obtained with various experiments. Finally, paper is concluded with summary of the work in section VII.

II. RELATED WORK

The simplest form of speech enhancement primitive is the noise reduction from the noisy speech and is applicable for single channel based speech applications. In this type of speech enhancement techniques, algorithms are either/combinely based on the model of noisy speech or/and perceptual model of speech using masking threshold. The generalized diagram of single channel enhancement technique is shown in fig. 2.

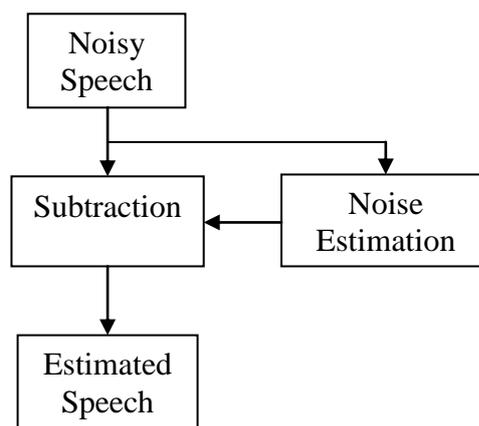


Fig. 2 Single Channel Speech Enhancement Technique

A. Spectral Subtraction Technique

Spectral subtraction technique is historically one of the first algorithms proposed for background noise reduction [5] as shown in Fig. 3 Spectral subtraction is performed by subtracting an estimate of the noise spectrum from the noisy speech spectrum as shown in Fig. 3.

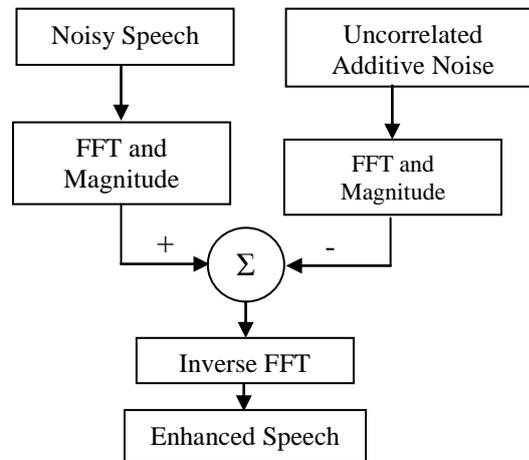


Fig. 3 Spectral Subtraction Speech Enhancement Method

That is

$$y(n) = x(n) + d(n) \quad (1)$$

Where $y(n)$, $x(n)$ and $d(n)$ denote the noisy speech, clean speech and uncorrelated additive noise, respectively. In frequency domain it can be written as

$$Y(k) = X(k) + D(k) \quad (2)$$

Where 'k' is index of frequency. Estimated clean speech spectrum is obtained as

$$|\hat{X}(k)| = |Y(k)| - |\hat{D}(k)| \quad (3)$$

Where $\hat{D}(k)$ is average magnitude of noise spectrum.

B. Wavelet Transform Technique

Spectral subtraction most widely used in single microphone algorithms for speech enhancement but it produces difficulties in pause detection due to additional relic as musical noise. DONOHA [6] in 1995, presented approach for denoising signal degraded by additive white noise using wavelet thresholding technique. Even though several papers [7, 8] were presented the application of wavelet transform for speech enhancement, but there are many issues yet to be determined for speech enhancement, where speech signals degraded by various noise sources.

In wavelet domain representation wavelet concentrate speech information into a few neighbouring coefficients. So as a result, taking wavelet transform of a speech signal, many coefficients will either be zero or have negligible magnitudes. Different steps involved in implementation of speech enhancement using wavelet transform are shown in Fig.4.

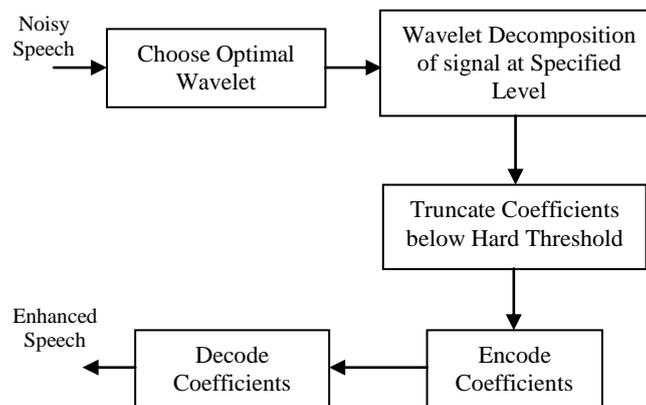


Fig. 4 Wavelet Transform based Speech Enhancement

Optimal wavelet can be selected by energy conservation properties. In this paper Daubechies wavelet (dB10) is used because it is observed that it concentrate more than 96% of signal energy in the level approximation coefficients. Next step is to decompose a signal into different resolutions or frequency bands, which are carried by choosing wavelet function and determining the Discrete Wavelet Transform (DWT). Here decomposition level chosen to be at scale 5 which is adequate for processing of speech signal. Next task is to truncating small value wavelet coefficients below a threshold, hard thresholding used in this experiment. After truncating, small value coefficients are encoded efficiently with two bytes one byte indicates a sequence of zeros in wavelet transforms vector and the second byte representing the number of consecutive zeros. Further synthesizing noisy speech signal using inverse discrete wavelet transform results in enhanced speech signal.

C. Adaptive Wiener Filtering

Wiener filtering estimates noise free speech signal from that noisy speech signal corrupted by additive noise. Estimation is performed by minimizing the Mean Square Error (MSE) between the noise free signal $x(n)$ and its estimation $\hat{x}(n)$. The problem with this method is that it has fixed frequency response at all frequencies and it also required estimation of power spectral densities of noise free and noise signal before filtering. To solve this problem, M.A. Abd E-Fattah [9] presented adaptive wiener filtering approach in 2008. According to this approach enhanced speech signal of small segment stationary noisy signal can be represented as

$$\hat{x}(n) = m_x + (x(n) - m_x) * \frac{V_x}{V_x + V_d} \quad (4)$$

Where m_x is mean of noise free speech signal, V_x and V_d are variance of noise free speech and noise respectively. If ' v_x ' is smaller than ' v_d ', input signal $x(n)$ is attenuated due to filtering effect. Different steps involved in implementation of speech enhancement using Adaptive Wiener Filtering are shown in Fig.5.

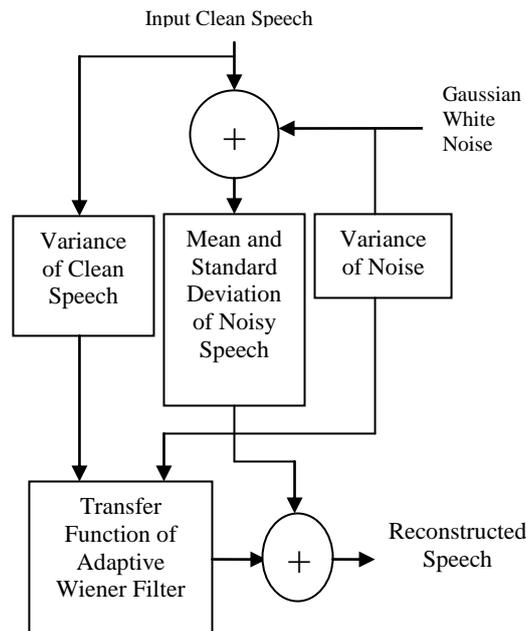


Fig.5. Adaptive Wiener Filtering based Speech Enhancement

III. OBJECTIVE EVALUATION

The objective comparison of three single channel speech enhancements is carried by evaluating performance of parameters such as, Mean Square Error (MSE), Normalized Mean Square Error (NRMSE), Signal to Noise Ratio (SNR), Peak Signal to

Noise Ratio (PSNR) and Average Absolute Distortion (AAD). It is based on mathematical comparison of the original and processed speech signals.

A. Signal to Noise Ratio (SNR)

It is most widely used and popular method to measure the quality of speech. It is ratio of signal to noise power in decibals.

$$\text{SNR (dB)} = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_d^2} \right) \quad (5)$$

Where σ_x^2 is the mean square of speech signal and σ_d^2 is the mean square difference between the original and reconstructed speech.

B. Peak Signal to Noise Ratio (PSNR)

$$\text{PSNR (dB)} = 10 \log_{10} \left(\frac{NX^2}{\|x-r\|^2} \right) \quad (6)$$

Where N is the length of the reconstructed signal, X is the maximum absolute square value of signal 'x' and $\|x-r\|^2$ is the energy of the difference between the original and reconstructed signal.

C. Normalized Root Mean Square Error (NRMSE)

$$\text{NRMSE} = \sqrt{\frac{[X(n)-r(n)]^2}{[x(n)-\mu x(n)]^2}} \quad (7)$$

D. Mean Square Error (MSE) and Average Absolute Distortion (AAD)

$$\text{MSE} = \frac{1}{N} [(r(n) - x(n))^2] \quad (8)$$

$$\text{AAD} = \frac{1}{N} |r(n) - x(n)| \quad (9)$$

Where N is length of input speech signal, x(n) is input speech signal and r(n) is reconstructed speech signal.

IV. SUBJECTIVE EVALUATION

The subjective comparison of three single channel speech enhancement techniques is measured to evaluate the speech quality along three dimensions: signal distortion (SIG) in this listener rate the speech signal alone using a five point scale of signal distortion, noise distortion (BAK) where listener rate the background noise alone using five point scale of background intrusiveness and overall quality (OVRL), where listener rate the overall effect using the scale of the Mean Opinion Score (MOS) [1=bad, 2=poor, 3=fair, 4=good, 5=excellent][10]. A total of 10 listeners were recruited for the listening test between the ages 20 to 40 years and the test lasted approximately one hours. Listeners given maximum score to Wavelet Transform based single channel speech enhancement technique due to its good speech reconstruction quality.

V. MATLAB GRAPHICAL USER INTERFACE IMPLEMENTATION (GUI)

For better understanding MATLAB Graphical User Interface (GUI) is developed for objective performance comparison of single channel speech enhancement techniques. GUIDE command of MATLAB is used to build this interface. In figure window for development of GUI Push Buttons, Static Text, Edit Text and Axes buttons are used. After putting all items in figure window, each item is programmed in MATLAB. Fig. 6 shows how all this should be seen.

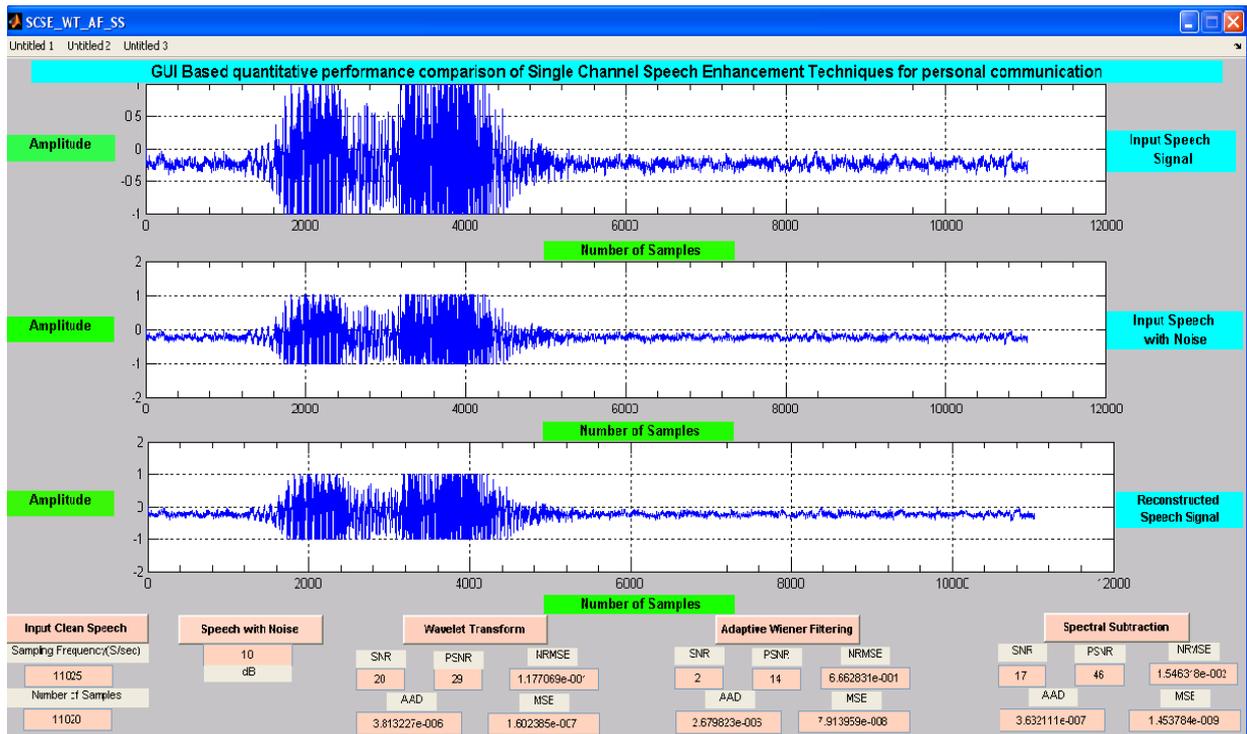


Fig.6 GUI for Comparison of Single Channel Speech Enhancement Techniques

VI. EXPERIMENTAL RESULTS

This section shows the effectiveness of our proposed work. The experimental results that concerned to our single channel speech enhancement systems were compared to Wavelet Transform (WT), Adaptive Wiener Filtering (AWF) and Spectral Subtraction (SS) methods. Test for speech enhancement were performed using uncontaminated recorded “Hello” word, which have 11020 samples, one second length, data size of 22040 bytes and PCM 11.025KHz, 16 bit Mono audio format using sound recorder of PC. This word is then contaminated with white gaussian noise type SNR of 0,-10,-20,-30,-40,-50 and -60dB to show the ability of single channel speech enhancement techniques for improving SNR in noisy speech environment for personal communication. In this paper the parameters used to compare performance of speech enhancement techniques are Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE), Normalized Mean Square Error (NRMSE) and Average Absolute Difference (AAD).

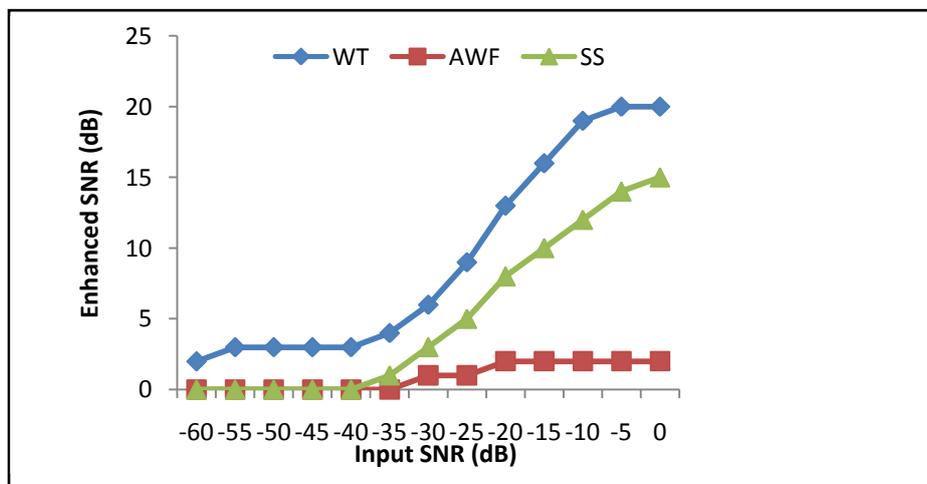


Fig. 7 Comparison of Enhanced SNR for WT, AWF and SS methods

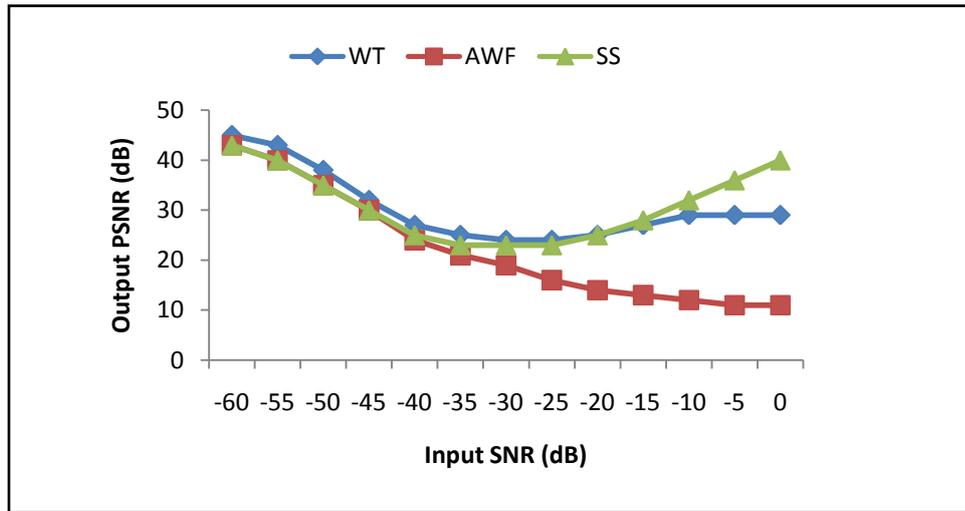


Fig. 8 Comparison of Output SNR for WT, AWF and SS methods

Fig.7 shows that WT method perform well throughout the input SNR range (-60 to 0 dB), while performance of AWF and SS remain almost same for the range -60 to -40 dB noise level. For the range -30 to 0 dB, SS method shows significant improvement in enhanced SNR value compare to AWF method. Fig. 8 shows PSNR for WT is slightly greater for the range from -60 to -25 dB compare to AWF and SS methods, while SS method shows significant improvement from -20 to 0 dB input SNR.

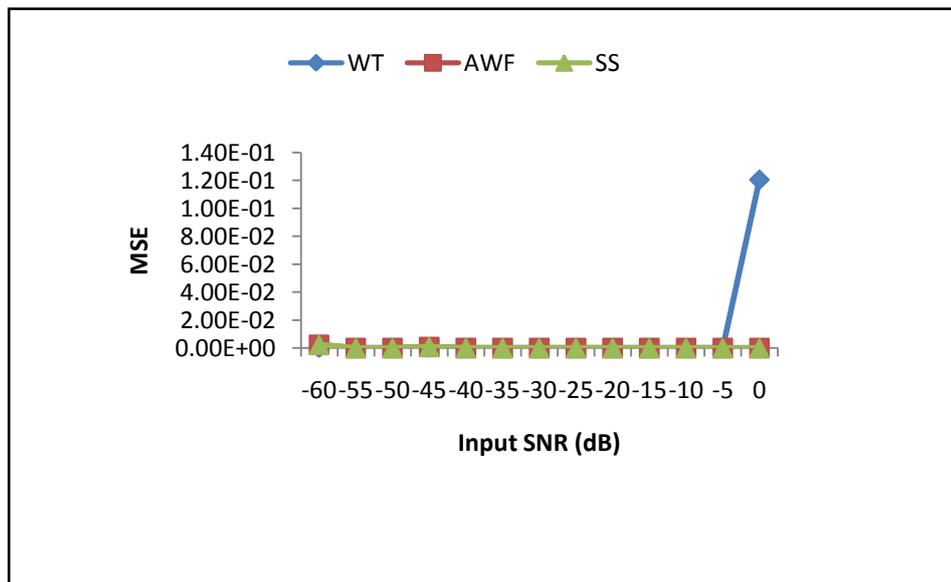


Fig. 9 Comparison of MSE for WT, AWF and SS methods

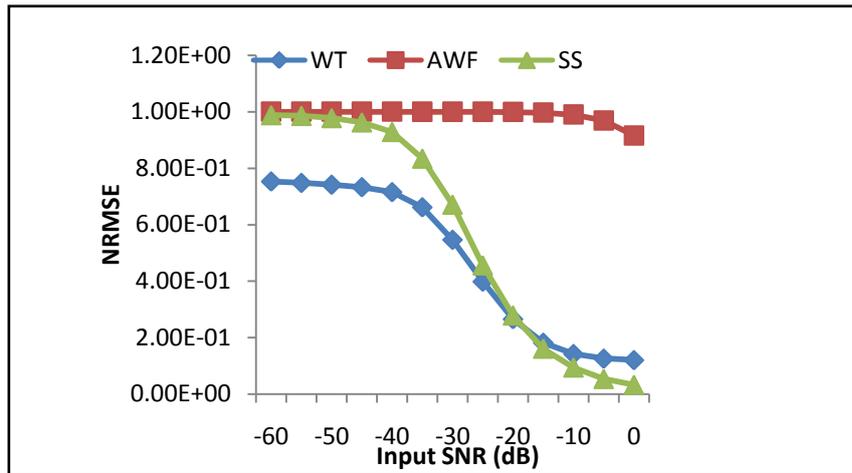


Fig. 10 Comparison of NRMSE for WT, AWF and SS methods

Fig. 9 shows that MSE is less for SS method compare to other methods, while Fig. 10 shows that NRMSE is less for WT compared to SS and AWF methods.

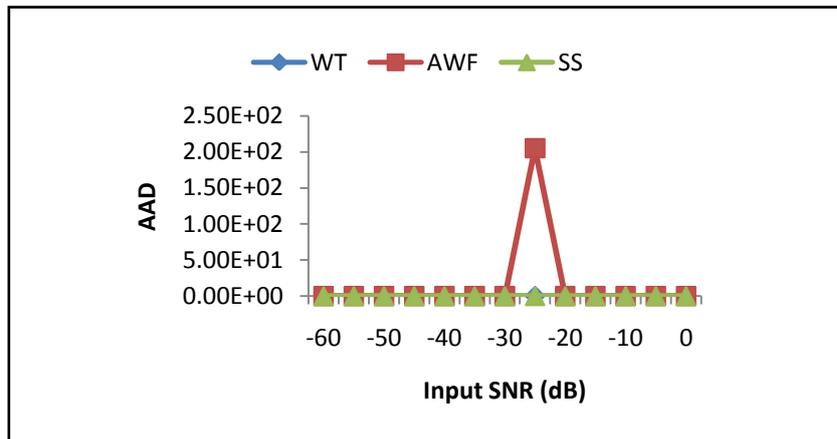


Fig. 11 Comparison of AAD for WT, AWF and SS methods

Fig. 11 shows that AAD for WT and SS is minimum for the range of -60dB to 0dB, while performance of AWF is poor for the range of -30 to -20 dB of input SNR.

TABLE I : Performance Measure of WT, SS and AWF based on Enhanced SNR

Input SNR (dB)	Enhanced SNR (dB)		
	WT	SS	AWF
0	20	15	2
-10	19	12	2
-20	13	8	2
-30	6	3	1
-40	3	0	0
-50	3	0	0
-60	2	0	0

TABLE II: Performance Measure of WT, SS and AWF based on Output PSNR

Input SNR (dB)	Output PSNR (dB)		
	WT	SS	AWF
0	29	40	11
-10	29	32	12
-20	25	25	14
-30	24	23	19
-40	27	25	24
-50	38	35	35
-60	45	43	43

TABLE III : Performance Measure of WT, SS and AWF based on Output MSE

Input SNR (dB)	Output MSE		
	WT	SS	AWF
0	1.2E-001	1.7E-008	1.1E-007
-10	2.2E-008	3.2E-008	3.6E-008
-20	8.3E-008	6.4E-008	1.2E-007
-30	2.0E-006	1.1E-005	1.0E-005
-40	9.0E-007	1.8E-005	1.3E-005
-50	1.1E-005	2.7E-005	8.1E-005
-60	2.6E-004	2.3E-003	2.3E-003

TABLE IV: Performance Measure of WT, SS and AWF based on Output NRMSE

Input SNR (dB)	Output NRMSE		
	WT	SS	AWF
0	1.2E-001	3.2E-002	9.1E-001
-10	1.4E-001	9.4E-002	9.9E-001
-20	2.6E-001	2.7E-001	9.9E-001
-30	5.4E-001	6.7E-001	9.9E-001
-40	7.1E-001	9.2E-001	9.9E-001
-50	7.4E-001	9.7E-001	9.9E-001
-60	7.5E-001	9.8E-001	9.9E-001

TABLE V: Performance Measure of WT, SS and AWF based on Output AAD

Input SNR (dB)	Output AAD		
	WT	SS	AWF
0	1.0E-006	1.2E-006	3.1E-006
-10	1.4E-006	1.7E-006	1.8E-006
-20	2.7E-006	2.4E-006	3.4E-006
-30	1.3E-005	3.2E-005	3.1E-005
-40	9.0E-007	4.1E-005	3.4E-005
-50	3.1E-005	4.9E-005	8.5E-005
-60	1.5E-004	4.6E-004	4.6E-004

From the above Table I to Table V results, it is clearly found that WT based single channel speech enhancement performed well in both clean and noisy environment.

VII. CONCLUSION

The main objective of the speech enhancement is to bring up the performance in the presence of noise and echo-interference to the performance obtained with pure speech signals, which is the ideal case. Thus, our aim was to approach the performance of single channel based speech enhancement techniques to that in the case of ideal signal. Another objective of this work is to compare objective and subjective performance of WT, AWF and SS based single channel speech enhancement techniques and the parameters used for comparison are Mean Square Error, Normalized Mean Square Error, Signal to Noise Ratio, Peak Signal to Noise Ratio and Average Absolute Distortion. Subjective evaluation also proved that Wavelet Transform based speech enhancement technique perform better due to good speech reconstruction quality. MATLAB GUI developed for speech enhancement techniques which help to be able to visualize the results obtained throughout this paper.

ACKNOWLEDGMENT

The authors thank Research Project of University of Mumbai, Mumbai, India for funding the project on “Quantitative Performance Comparison of Single Channel Speech Enhancement Techniques for Personal Communication”.

REFERENCES

- [1] V. Radha, Vimala C. and M. Krrishnaveni “Comparative Analysis of Compression Techniques for Tamil Speech Datasets”, IEEE-International Conference on Recent Trends in Information Technology (ICRITIT), 978-1-4577-0590-8/11/\$26.00 © 2011 IEEE, 2011.
- [2] P.M. Halder and A. K. M. Fazlul Haque, “Improved Echo Cancellation in VOIP”, International Journal of Advanced Computer Science and Applications(IJACSA), Vol. 2, No. 11, pp. 122-125, 2011.
- [3] A.M. Kondozi, Digital Speech Coding for Low Bit Rate Communication Systems, John Wiley and Sons. Ltd., pp 379-405, 2007.
- [4] Milind U. Nemade and Prof. Satish K. Shah, “Improvement in Speech Recognition Performance using Beamforming based Speech Enhancement,” International Journal of Electronics Communicationand Computer Engineering (IJECCCE), Vol.3, Issue 4, pp.745-751 , 2012
- [5] Boll S. F., "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust. Speech, Signal Processing, vol.ASSP-27, pp.113-120, 1979.
- [6] Donoha D.L., "Denoising by soft thesholding," IEEE Transactions on Information Theory, vol.41, no.3, pp. 613- 627, 1995.
- [7] Subhra Debdas, Vaishali Jagrit, Chinmay Chandrakar, and M.F. Quereshi, "Application of wavelet transform for speech processing," International Journal of Engineering Science and Technology (IJEST), Vol.3, No.8, pp.6666-6670, 2011.
- [8] C. Gargour, M. Gabrea, V. Ramachandran, and J. Lina, "A Short Introduction to Wavelets and Their Applications," IEEE Circuits and Systems Magazine, ISSN: 1531-636X, Vol.2, pp. 57-67, 2009.
- [9] M.A. Abd E-Fattah, M. I. Dessouky, S. M. Diab and F. E. Abd El-samie, "Speech Enhancement using Adaptive Wiener Filtering Approach," Progress in Electromagnetic Research M, Vol.4, pp.167-184, 2008.
- [10] D. Deepa, C. Prakarsha, and A. Shanmugam, "Single Channel Speech Enhancement using Spectral Gain Shaping method and Dereverberation for Digital Hearing Aid," International Conference on Computer Communication and Informatics(ICCCI), 978-1-4577-1583-9/12/\$26.00© 2012 IEEE, Coimbatore, India, 2012.

BIOGRAPHY



Milind U. Nemade was born in Maharashtra, India 1974. He graduated from the Amaravati University, Maharashtra, India in 1995. He received M.E (Electrical) degree with specialization in Microprocessor Applications from M.S. University of Baroda, Gujrat, India in 1999. Now he is Associate Professor and Head at Department of Electronics and Telecommunication, K.J. Somaiya Institute of Engineering and Information Technology Sion, Mumbai, University of Mumbai, India. He started PhD study at Electrical Department, Faculty of Technology and Engineering, M.S. University of Baroda, Gujrat, India. He presented and published four papers in national conferences and three papers in the proceedings of international conferences. His research interest includes speech and audio processing.



Prof. Satish K. Shah is a professor in the Electrical Engineering Department at Faculty of Technology, MS University of Baroda for **last Twenty Five** years. He is a fellow of IE(I) for past Fourteen Years and has also served as the member of Committee of Vadodara Local Center for more than SIX years in past. He has guided more than hundred projects at UG/PG level and completed a research project on DSP based Active Power filter sponsored by AICTE, New Delhi. He has written three books on Embedded System design/Microprocessors/Microcontrollers and presented/published more than 25 research papers in national/international conferences/Journals. He has attended and organized several seminars, workshops, and symposiums for UGC, AICTE, IETE, and MSU. He is a fellow of other technical associations such as: IETE, ISA and IEEE (NY) & ISTE. He has served as the member, Hon Secretary and Treasurer of their local executive committees for a span of six-eight years